11/21 Recitation

1. Consider the vectors $\mathbf{x}_1 = \begin{bmatrix} 3 \\ 0 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \mathbf{x}_3 = \begin{bmatrix} 0 \\ 3 \end{bmatrix} \in \mathbb{R}^2$. Let $A = \begin{bmatrix} 3 & 0 \\ 1 & 1 \\ 0 & 3 \end{bmatrix}$.

   Consider the following SVD of A

$$\begin{bmatrix} 3 & 0 \\ 1 & 1 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 3\sqrt{22}/22 & -\sqrt{2}/2 & \sqrt{11}/11 \\ \sqrt{22}/11 & 0 & -3\sqrt{11}/11 \\ 3\sqrt{22}/22 & \sqrt{2}/2 & \sqrt{11}/11 \end{bmatrix} \begin{bmatrix} \sqrt{11} & 0 \\ 0 & 3 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix}^T$$

   (a) Find and sketch the best fit $1-$dimensional subspace of $\mathbf{x}_1, \mathbf{x}_2$, and $\mathbf{x}_3$. What quantity does this line minimize?

   (b) Find and sketch the least squares line of best fit of $\mathbf{x}_1, \mathbf{x}_2$, and $\mathbf{x}_3$. What quantity does this line minimize?

   (c) Which of the two lines gives a better approximation of the data? What feature of the line found in part (b) precludes it from being the best fit $1-$dimensional subspace?

(d) **Looking Ahead**: If we *center* the data about the origin, then the best fit $1-$dimensional subspace will provide a more accurate line of best fit. This is what is done in *principle component analysis* (which we will discuss next week). Let's do this in this example. Let A' be the matrix obtained by subtracting each entry of the ith column of A by the average of the entries in the ith column of A. This yields the matrix

$$A' = \begin{bmatrix} 5/3 & -4/3 \\ -1/3 & -1/3 \\ -4/3 & 5/3 \end{bmatrix}$$

Plot the related data points in $\mathbb{R}^2$ and compare it to the old data points; notice that they look like the old data points, but shifted towards the origin. Next, use Julia to find the $1-$dimensional subspace of best fit; sketch it.

2. Best fit subspaces are not necessarily unique. Consider the following data points in $\mathbb{R}^3$.

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \mathbf{x}_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

(a) Find an SVD of the matrix A whose rows are $\mathbf{x}_i^T$ (by hand).

(b) Determine (at least) two best fit 2-dimensional subspaces.

(c) Visit `https://www.geogebra.org/m/qsfprdt3` to see two such $2-$dimensional best fit subspaces. Determine $\sum ||e_i||^2$ for each plane (geometrically) and verify that they give the same value.

3. Write a Julia function that takes in vectors of the same dimension, along with an integer k, and outputs a basis for the best fit $k-$dimensional subspace of those vectors.